

Envelope Word and Gap Sequence in Doubling Sequence

Yuke Huang^{1,2} Hanxiong Zhang^{3,4}

ABSTRACT

Let ω be a factor of Doubling sequence $D_\infty = x_1x_2\cdots$, then it occurs in the sequence infinitely many times. Let ω_p be the p -th occurrence of ω and $G_p(\omega)$ be the gap between ω_p and ω_{p+1} . In this paper, we discuss the structure of the gap sequence $\{G_p(\omega)\}_{p \geq 1}$. We prove that all factors can be divided into two types, one type has exactly two distinct gaps $G_1(\omega)$ and $G_2(\omega)$, the other type has exactly three distinct gaps $G_1(\omega)$, $G_2(\omega)$ and $G_4(\omega)$. We determine the expressions of gaps completely. And also give the substitution of each gap sequence. The main tool in this paper is “envelope word”, which is a new notion, denoted by $E_{m,i}$. As an application, we determine the positions of all ω_p , discuss some combinatorial properties of factors, and count the distinct squares beginning in $D_\infty[1, N]$ for $N \geq 1$.

Key Word: Envelope word, Gap sequence, Doubling sequence, Position, Square.

1. Introduction

Factor property have been studied extensively, such as Lothaire[9, 10]. Doubling sequence, sometimes appears in a different name: period-doubling sequence, is a classical example of a substitution of constant length over a binary alphabet. It is also the difference of Thue-Morse sequence [3, 12]. The Doubling sequence, among other so-called primitive substitution sequence, has many remarkable properties, and it appears in many aspects of mathematics and computer science, symbolic dynamics, theoretical computer science etc., we refer to Allouche and Shallit[1], Berstel[2] and also Mauduit[11].

Wen and Wen[13] studied the factor structure of Fibonacci sequence, where they defined the singular word and give the positively separate property of the singular words. Huang and Wen[8] extend the results from singular words to arbitrary words ω of the Fibonacci sequence, and discuss the structure of gap sequence $\{G_p(\omega)\}_{p \geq 1}$. The main aim of this article is to extend the results in Huang and Wen[8] from Fibonacci sequence to Doubling sequences D_∞ .

The main result in this paper is as follows.

Theorem (Gap sequence of factor $\omega \prec D_\infty$).

- (1) All factors can be divided into two types;
- (2) One type of factor has exactly two distinct gaps $G_1(\omega)$ and $G_2(\omega)$, the gap sequence is $\varphi_1(D_\infty)$, where $\varphi_1(a, b) = (a, bb)$;
- (3) The other type of factor has exactly three distinct gaps $G_1(\omega)$, $G_2(\omega)$ and $G_4(\omega)$, the gap sequence is $\varphi_2(D_\infty)$, where $\varphi_2(a, b) = (ab, acac)$.

The main tool in this paper is “envelope word”. Using it, we can determine the type of each factor, and give the expressions of each gap $G_p(\omega)$ and each substitution φ_i . Then we can determine the structure of gap sequence of D_∞ completely.

This paper is organized as follows.

Section 1 is devoted to the introduction and preliminaries. In Section 2, we give some basic

¹Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, P. R. China.

²E-mail address: hyg03ster@163.com.

³Department of Mathematics, China University of Mining and Technology, Beijing, 100083, P. R. China.

⁴E-mail address: zhanghanxiong@163.com(Corresponding author).

properties of Doubling sequence. In Section 3, we define a new notion “envelope word” and give the weak version of “uniqueness of envelope extension property”, there are two types of envelope words. In Section 4, we discuss the gaps and gap sequence of envelope words. Then we give the strong version of “uniqueness of envelope extension property” in Section 5, which makes envelope words quite special. Using them, we determine the gaps and gap sequence of an arbitrary word. As an application, we determine the positions of ω_p for all (ω, p) in Section 6. In Section 7, we study some combinatorial properties of factors in Doubling sequence.

1.1 Notation and Basic Properties

Let $\mathcal{A} = \{a, b\}$ be a binary alphabet. Let \mathcal{A}^* be the set of finite words on \mathcal{A} and $\mathcal{A}^{\mathbb{N}}$ be the set of one-sided infinite words. The elements of \mathcal{A}^* are called words or factors, which will be denoted by ω . The neutral element of \mathcal{A}^* is called the empty word, which we denote by ε . For a finite word $\omega = x_1x_2 \cdots x_n$, the length of ω is equal to n and denoted by $|\omega|$. The number of occurrences of letter $\alpha \in \mathcal{A}$ in ω is denoted by $|\omega|_{\alpha}$.

The concatenation of two words $\nu = x_1x_2 \cdots x_r$ and $\omega = y_1y_2 \cdots y_n$ is the word $\nu\omega = x_1x_2 \cdots x_ry_1y_2 \cdots y_n$. This operation is associative and has a unit element, the empty word ε .

As usual, the Doubling sequence D_{∞} is defined as the fixed point beginning with letter a under the Doubling substitution σ , which is defined over the alphabet $\mathcal{A} = \{a, b\}$ by $\sigma(a) = ab$ and $\sigma(b) = aa$:

$$D_{\infty} = \sigma^{\infty}(a) = abaaabababaaabaaababababaaabab \cdots$$

We also use the notations that $A_m = \sigma^m(a)$ and $B_m = \sigma^m(b)$ where $m \geq 0$. Obviously, $|A_m| = |B_m| = 2^m$.

The notation $\nu \prec \omega$ means that word ν is a factor of word ω .

Let $\tau = x_1x_2 \cdots x_m \cdots$ be a sequence. For any $i \leq j$, define $\tau[i, j] := x_i x_{i+1} \cdots x_{j-1} x_j$, the factor of τ of length $j - i + 1$, starting from the i -th letter and ending to the j -th letter. By convention, we note $\tau[i] := \tau[i, i] = x_i$ and $\tau[i, i-1] := \varepsilon$.

If $\omega = \tau[i, j]$, τ is a word or a sequence, ω is said to be occur at position i in τ . Furthermore, $L(\omega, p)$ denotes the position of ω_p .

Let $\nu = \nu_1\nu_2 \cdots \nu_n \in \mathcal{A}^*$, we denote by $\nu^{-1} := \nu_n^{-1} \cdots \nu_2^{-1} \nu_1^{-1}$, called the inverse word of ν . If $\omega = u\nu$, then $\omega^{-1} = (u\nu)^{-1} = \nu^{-1}u^{-1}$. Furthermore, $\omega\nu^{-1} = u\nu\nu^{-1} = u$ and $u^{-1}\omega = u^{-1}u\nu = \nu$.

1.2 Factor Sequence and Gap Sequence

Let ω be a factor of Doubling sequence D_{∞} . In this subsection, we will introduce some definitions: factor sequence $\{\omega_p\}_{p \geq 1}$, gap word $G_p(\omega)$, gap sequence $\{G_p(\omega)\}_{p \geq 1}$, etc.

Definition 1.1 (Factor sequence). *Let ω be a factor of Doubling sequence, then it occurs in the sequence infinitely many times, which we arrange by the sequence $\{\omega_p\}_{p \geq 1}$, where ω_p denote the p -th occurrence of ω .*

Definition 1.2 (Gap). *Let $\omega_p = x_{i+1} \cdots x_{i+n}$, $\omega_{p+1} = x_{j+1} \cdots x_{j+n}$, the gap between ω_p and ω_{p+1} , denoted by $G_p(\omega)$, is defined by*

$$G_p(\omega) = \begin{cases} \varepsilon & \text{when } i+n = j, \omega_p \text{ and } \omega_{p+1} \text{ are adjacent;} \\ x_{i+n+1} \cdots x_j & \text{when } i+n < j, \omega_p \text{ and } \omega_{p+1} \text{ are separated;} \\ (x_{j+1} \cdots x_{i+n})^{-1} & \text{when } i+n > j, \omega_p \text{ and } \omega_{p+1} \text{ are overlapped.} \end{cases}$$

The set of gaps of factor ω is defined as $\{G_p(\omega) \mid p \geq 1\}$.

Example. $G_1(aa) = a^{-1}$ (overlapped), $G_2(aa) = babab$ (separated), $G_2(aab) = \varepsilon$ (adjacent).

Remark. A closely related concept of “Gap” is “Return Word”, introduced by F.Durand[7], for characterizing a sequence over a finite alphabet to be substitutive, he proved that a sequence is primitive substitutive if and only if the set of its return words is finite, which means, each factor of this sequence has finite return words. In 2001, L.Vuillon[14] proved that an infinite word τ is a Sturmian sequence if and only if each non-empty factor $\omega \prec \tau$ has exactly two distinct return words. Some other related researches (see also [5]) focused in the cardinality of the set of return words of ω and the consequent results, but didn't concern about the structures of the sequences derived by return words.

Essentially gap words can be derived from the return words by deleting one prefix ω , but since the terminology “gap” will be convenient and have some advantages for our discussions, we prefer to adopt it.

Definition 1.3 (Gap sequence). *Let $G_p(\omega)$ be the gap between ω_p and ω_{p+1} , we call $\{G_p(\omega)\}_{p \geq 1}$ the gap sequence of the factor ω .*

2. Some Basic Properties of Doubling Sequence

In this section, we will give some very important properties of Doubling sequence D_∞ .

Definition 2.1 (δ_m). *Let δ_m be the last letter of A_m .*

Property 2.2. (1) δ_{m-1} be the last letter of B_m ;
(2) For $m \geq 0$, $\delta_m = a$ (resp. b) when m is even (resp. odd).

Proposition 2.3. $A_m \delta_m^{-1} = B_m \delta_{m+1}^{-1}$ for $m \geq 0$.

Property 2.4 (Palindrome). $A_m \delta_m^{-1}$ is a palindrome for $m \geq 0$.

Proof. By induction, when $m = 0, 1, 2$, $A_0 \delta_0^{-1} = \varepsilon$, $A_1 \delta_1^{-1} = a$, $A_2 \delta_2^{-1} = aba$. Assume $A_m \delta_m^{-1}$ is a palindrome. Since $A_{m+2} \delta_{m+1}^{-1} = A_m B_m A_m A_m \delta_m^{-1}$ and $B_m \delta_{m-1}^{-1} = A_m \delta_m^{-1}$,

$$A_{m+2} \delta_{m+1}^{-1} = A_m \delta_m^{-1} * \delta_m * A_m \delta_m^{-1} * \delta_{m-1} * A_m \delta_m^{-1} * \delta_m * A_m \delta_m^{-1}.$$

That means $A_{m+2} \delta_{m+1}^{-1}$ is a palindrome too. □

Lemma 2.5 (Positions). *For $m \geq 1$,*

- (1) A_m occurs in $A_m B_m A_m$ twice at positions 1 and $2^{m+1} + 1$;
- (2) A_m occurs in $A_m A_m$ twice at positions 1 and $2^m + 1$;
- (3) B_m occurs in $B_m B_{m-1} B_{m+1} B_m$ twice at positions 1 and $7 * 2^{m-1} + 1$;
- (4) B_m occurs in $B_m B_{m-1} B_m$ twice at positions 1 and $3 * 2^{m-1} + 1$.

These properties above will be used in the proofs of Theorem 4.1, 4.2 and 4.3.

3. Envelope Word $E_{m,i}$

In this section, we will discuss a particular kind of factor, called envelope word, denoted by $E_{m,i}$. We will introduce some definitions: envelope word $E_{m,i}$, envelope set \mathcal{E} , envelope word of factor ω , denoted by $Env(\omega)$, etc.

Definition 3.1 (Envelope set and envelope word in Doubling sequence).

Envelope set $\mathcal{E} := \{E_{m,i}, m \geq 1, i = 1, 2\}$, where

$$E_{m,1} = A_m \delta_m^{-1}, \quad E_{m,2} = A_{m-1} A_m \delta_m^{-1}.$$

The element $E_{m,i}$ in \mathcal{E} is called an envelope word in Doubling sequence.

Property 3.2. For $m \geq 1$ and $i = 1, 2$, $E_{m+1,i} = \sigma(E_{m,i})a$.

Example. $E_{1,1} = a$, $E_{2,1} = aba$, $E_{3,1} = abaaaba$; $E_{1,2} = aa$, $E_{2,2} = ababa$, $E_{3,2} = abaaabaaaba$.

Property 3.3. For $m \geq 1$, $E_{m+1,1} = E_{m,1} \delta_m E_{m,1}$ and $E_{m+1,2} = E_{m,1} \delta_m E_{m,1} \delta_m E_{m,1}$.

Property 3.4 (Palindrome). Envelope word $E_{m,i}$ is a palindrome for $m \geq 1$, $i = 1, 2$.

Definition 3.5 (Order in Envelope Set \mathcal{E}).

Let E_{m_1,i_1} and E_{m_2,i_2} be two elements in envelope set \mathcal{E} .

- (1) if $m_1 < m_2$ then $E_{m_1,i_1} \sqsubset E_{m_2,i_2}$;
- (2) if $m_1 = m_2$ and $i_1 < i_2$ then $E_{m_1,i_1} \sqsubset E_{m_2,i_2}$.

Example. $E_{1,1} \sqsubset E_{1,2} \sqsubset E_{2,1}$.

Definition 3.6 (Envelope word of factor ω).

The envelope word of factor $\omega \in D_\infty$, denoted by $Env(\omega) := E_{m,i}$, is the minimal element in envelope set \mathcal{E} under the order " \sqsubset " subject to $\omega \prec E_{m,i}$.

Example. $Env(bab) = ababa = E_{2,2}$, $Env(aaa) = abaaaba = E_{3,1}$.

Theorem 3.7 (Uniqueness of envelope extension, weak version).

Let $Env(\omega) = E_{m,i}$, $i = 1, 2$, then ω has an unique envelope extension as

$$E_{m,i} = \mu_1(\omega) * \omega * \mu_2(\omega),$$

where $\mu_1(\omega)$ and $\mu_2(\omega)$ are constant words depending only on ω .

Proof. **Part 1.** When $i = 1$, $E_{m,1} = A_m \delta_m^{-1}$. It contains $E_{m-1,1}$ and $E_{m-2,2}$ as subwords.

$$\begin{array}{c}
 \begin{array}{c}
 v_2 = \delta_{m-1} A_{m-2} \\
 \hline
 v_1 = \delta_{m-1} A_{m-3} \\
 \hline
 A_{m-3} A_{m-2} \delta_{m-2}^{-1} = E_{m-2,2} \\
 \hline
 \end{array} \\
 E_{m,1} = \overbrace{A_{m-2} A_{m-3} \delta_{m-1}^{-1} \delta_{m-1} A_{m-3} \delta_{m-1}^{-1} \delta_{m-1} A_{m-2} \delta_{m-1}^{-1} \delta_{m-1} A_{m-2} \delta_{m-1}^{-1} \delta_{m-1} A_{m-2} \delta_{m-1}^{-1} \delta_{m-1} A_{m-2} \delta_{m-1}^{-1}}^{A_{m-1} \delta_{m-1}^{-1} = E_{m-1,1} \quad A_{m-1} \delta_{m-1}^{-1} = E_{m-1,1}}
 \end{array}$$

Fig. 3.1: The $E_{m-1,1}$ and $E_{m-2,2}$ in $E_{m,1}$.

By the definition of $Env(\omega) = E_{m,1}$, ω must contain $v_1 = \delta_{m-1} A_{m-3}$ or $v_2 = \delta_{m-1} A_{m-2}$ as a subword, i.e. $\delta_{m-1} A_{m-3} \prec \omega$.

$$E_{m,1} = A_{m-3} B_{m-3} A_{m-3} A_{m-3} A_{m-3} B_{m-3} A_{m-3} A_{m-3} \delta_{m-3}^{-1},$$

so all possible positions of $\delta_{m-3} A_{m-3}$ in $E_{m,1}$ is $3 * 2^{m-3}$ and 2^{m-1} .

Suppose factor ω occurs twice in $E_{m,1}$, then factor $\delta_{m-3} A_{m-3}$ in ω occurs at the two positions above. Since the two ω are equal, consider the structure around the two $\delta_{m-3} A_{m-3}$:

$$E_{m,1} = A_{m-2} A_{m-3} \delta_{m-1}^{-1} \overbrace{A_{m-3} \delta_{m-1}^{-1} \delta_{m-1} A_{m-3} \delta_{m-1}^{-1} \delta_{m-1} A_{m-3} \delta_{m-1}^{-1}}^{v_1 = \delta_{m-1} A_{m-3}} \delta_{m-3} A_{m-3} \delta_{m-3}^{-1} \delta_{m-2} A_{m-2} \delta_{m-2}^{-1}$$

Fig. 3.2: The structure around the two $\delta_{m-3}A_{m-3}$.

So we get that

$$\omega \prec A_{m-3}A_{m-3}A_{m-3}\delta_{m-3}^{-1} = A_{m-3}A_{m-2}\delta_{m-2}^{-1} = E_{m-2,2} \sqsubset E_{m,1}.$$

It contradicts the definition of $Env(\omega) = E_{m,1}$.

This means ω occurs in $E_{m,1}$ only once, i.e. the envelope extension is unique.

Part 2. When $i = 2$, $E_{m,2} = A_{m-1}A_m\delta_m^{-1}$. It contains $E_{m,1}$ as subwords.

$$E_{m,2} = \underbrace{A_{m-1} \delta_{m-1}^{-1} \delta_{m-1} A_{m-1} \delta_{m-1}^{-1}}_{A_m \delta_m^{-1} = E_{m,1}} \delta_{m-1} B_{m-1} \delta_m^{-1}$$

Fig. 3.3: The $E_{m,1}$ occurs in $E_{m,2}$.

By the definition of $Env(\omega) = E_{m,2}$, ω must contain $\delta_{m-1}A_{m-3}$ or $v_2 = \delta_{m-1}A_{m-1}$ as a subword, i.e. $\delta_{m-1}A_{m-1} \prec \omega$.

$$E_{m,2} = A_{m-2}B_{m-2}A_{m-2}B_{m-2}A_{m-2}A_{m-2}\delta_m^{-1}.$$

So all possible positions of $\delta_{m-1} * A_{m-1}$ in $E_{m,1}$ is 2^{m-1} .

This means ω occurs in $E_{m,2}$ only once, i.e. the envelope extension is unique.

Lemma 3.8 (Positions). *For $m \geq 1$,*

- (1) $E_{m,1}$ occurs in $E_{m+1,1}$ twice at positions 1 and $2^m + 1$;
- (2) $E_{m,1}$ occurs in $E_{m,2}$ twice at positions 1 and $2^{m-1} + 1$;
- (3) $E_{m,2}$ occurs in $E_{m+2,1}$ twice at positions $2^m + 1$ and $3 * 2^{m-1} + 1$;
- (4) $E_{m,2}$ occurs in $E_{m+2,2}$ four times: $2^m + 1$, $3 * 2^{m-1} + 1$, $3 * 2^m + 1$ and $7 * 2^{m-1} + 1$.

Lemma 3.9. *For $m \geq 1$, $i = 1, 2$,*

- $$\begin{aligned} (1) \quad & E_{m,1} \prec E_{m',i}, \forall m' \geq m; \\ (2) \quad & E_{m,2} \not\prec E_{m+1,i}; \\ (3) \quad & E_{m,2} \prec E_{m'+2,i}, \forall m' \geq m. \end{aligned}$$

Property 3.10 (Envelope set \mathcal{E} is minimum).

Envelope set \mathcal{E} is minimum to ensure Theorem 3.7.

Proof. Let \mathcal{E}' be a proper subset of \mathcal{E} . $E_{m,i}$ is the minimal element of $\mathcal{E} - \mathcal{E}'$. Let

$$\mathcal{C} = \{E_{n,j} \in \mathcal{E}' : E_{m,i} \prec E_{n,j}\}.$$

If $\mathcal{C} = \emptyset$, $E_{m,i}$ does not have envelope word. It contradicts the definition of envelope word.

If $\mathcal{C} \neq \emptyset$, $Env(E_{m,i}) = \min\{E_{n,j} \in \mathcal{E}' : E_{m,i} \prec E_{n,j}\}$, denoted by $E_{m',i'}$.

Case 1: $i = 1$ and $m' > m$. By Lemma 3.9(1), $E_{m+1,1} \prec E_{m',i'}$. By Lemma 3.8(1), $E_{m,1}$ occurs in $E_{m+1,1}$ twice. So $E_{m,1}$ occurs in $Env(E_{m,1})$ more than once.

Case 2: $i = 1$ and $m' = m$, then $E_{m',i'} = E_{m,2}$. By Lemma 3.8(2), $E_{m,1}$ occurs in $Env(E_{m,1})$ more than once.

Case 3: $i = 2$. By Lemma 3.9(2), $m' > m + 1$. By Lemma 3.9(1), $E_{m+2,1} \prec E_{m',i'}$. By Lemma 3.8(3), $E_{m,2}$ occurs in $E_{m+2,1}$ twice. So $E_{m,2}$ occurs in $Env(E_{m,1})$ more than once.

All cases contradict the weak version of “uniqueness of envelope extension” in Theorem 3.7. So envelope set \mathcal{E} is minimum to ensure Theorem 3.7. \square

4. Gaps and Gap Sequence of Envelope Word $E_{m,i}$

Each envelope word $E_{m,i}$ occurs in Doubling sequence infinitely many times. Let $E_{m,i,p}$ be the p -th occurrence of $E_{m,i}$, $G_p(E_{m,i})$ be the gap between $E_{m,i,p}$ and $E_{m,i,p+1}$, $G_0(E_{m,i})$ be the prefix of D_∞ before $E_{m,i,1}$.

Theorem 4.1. *Let $G_0(E_{m,i})$ be the prefix of D_∞ before $E_{m,i,1}$, then*

- (1) $G_0(E_{m,1}) = D_\infty[1, 0] = \varepsilon$;
- (2) $G_0(E_{m,2}) = D_\infty[1, 2^m] = A_m$.

Proof. (1) Since $D_\infty[1, 2^m - 1] = A_m \delta_m^{-1} = E_{m,1}$, $G_0(E_{m,1}) = \varepsilon$.

(2) Since $D_\infty[1, 5 * 2^{m-1} - 1] = A_{m-1} B_{m-1} A_{m-1} A_{m-1} A_{m-1} \delta_{m+1} = A_m A_{m-1} B_m \delta_{m-1}$. By Proposition 2.3, $B_m \delta_{m+1} = A_m \delta_m$, so $D_\infty[2^m + 1, 5 * 2^{m-1} - 1] = A_{m-1} A_m \delta_m^{-1} = E_{m,2}$.

Consider the first A_{m-1} in $A_{m-1} A_m \delta_m^{-1}$. By Lemma 2.5(1), it occurs in $A_{m-1} B_{m-1} A_{m-1}$ twice at positions 1 and $2^m + 1$. Suppose this A_{m-1} occurs at positions 1, then the 2^m -th letter in $A_{m-1} B_{m-1} A_{m-1}$ is δ_m . But the letter in $A_{m-1} A_m \delta_m^{-1} = A_{m-1} A_{m-1} B_{m-1} \delta_m^{-1}$ at the same position is δ_{m-1} . It contradicts $\delta_m \neq \delta_{m-1}$.

This means $D_\infty[2^m + 1, 5 * 2^{m-1} - 1] = E_{m,2,1}$ and $G_0(E_{m,2}) = A_m$. \square

Theorem 4.2 (The first two gaps of $E_{m,1}$). *For $m \geq 1$,*

- (1) $G_1(E_{m,1}) = \delta_m$;
- (2) $G_2(E_{m,1}) = \delta_{m-1} A_{m-1}^{-1}$.

Proof. By the definition of envelope word, $E_{m,1} = A_m \delta_m^{-1}$. By Property 2.3, $B_m \delta_{m+1}^{-1} = A_m \delta_m$,

$$D_\infty[1, 2^{m+1} - 1] = A_m B_m \delta_{m-1}^{-1} = A_m A_m \delta_m^{-1} = A_m \delta_m^{-1} * \delta_m * A_m \delta_m^{-1}.$$

This means $D_\infty[2^m + 1, 2^{m+1} - 1] = E_{m,1}$.

Consider the first A_{m-1} in $E_{m,1} = A_m \delta_m^{-1} = A_{m-1} B_{m-1} \delta_m^{-1}$. By Lemma 2.5(1), it occurs in $D_\infty[1, 3 * 2^m] = A_{m-1} B_{m-1} A_{m-1}$ twice at positions 1 and $2^m + 1$. When the A_{m-1} occurs at positions 1, $E_{m,1} = D_\infty[1, 2^m - 1]$. By the proof of Theorem 4.1(1), it is $E_{m,1,1}$, i.e.,

$$E_{m,1,1} = D_\infty[1, 2^m - 1].$$

When the A_{m-1} occurs at positions $2^m + 1$, $E_{m,1} = D_\infty[2^m + 1, 2^{m+1} - 1]$, it is $E_{m,1,2}$, i.e.,

$$E_{m,1,2} = D_\infty[2^m + 1, 2^{m+1} - 1].$$

Using similar arguments and by Lemma 2.5(2), we can determine the positions of $E_{m,1,3}$,

$$E_{m,1,3} = D_\infty[3 * 2^{m-1} + 1, 5 * 2^{m-1} - 1].$$

This means:

$$\begin{aligned} G_1(E_{m,1}) &= D_\infty[2^m] = \delta_m; \\ G_2(E_{m,1}) &= D_\infty[3 * 2^{m-1} + 1, 2^{m+1} - 1]^{-1} = \delta_{m-1} A_{m-1}^{-1}. \end{aligned}$$

\square

Proof. We write $G_1(E_{m,2})$ (resp. $G_2(E_{m,2})$, $G_4(E_{m,2})$) as G_1 (resp. G_2 , G_4) for short.

Step 1. $E_{m,2} = A_{m-1}A_m\delta_m^{-1}$, $G_1 = \delta_m A_m^{-1}$, $G_2 = \delta_m B_{m+1}$ and $G_4 = \delta_m$.

(a) $E_{m,2}G_1E_{m,2}G_2 = A_{m-1}A_m\delta_m^{-1} * \delta_m A_m^{-1} * A_{m-1}A_m\delta_m^{-1} * \delta_m B_{m+1} = B_m A_m B_{m+1}$;

(b) $E_{m,2}G_1E_{m,2}G_4 = A_{m-1}A_m\delta_m^{-1} * \delta_m A_m^{-1} * A_{m-1}A_m\delta_m^{-1} * \delta_m = B_m A_m$;

Step 2. By the definition of sequence Θ_2 , $\sigma(ab, acac) = (abacac, abab)$. So we only need to prove the two properties below, where $\eta = B_m$ is a shift.

(1) $\eta\sigma(E_{m,2}G_1E_{m,2}G_2)\eta^{-1} = E_{m,2}G_1E_{m,2}G_2E_{m,2}G_1E_{m,2}G_4E_{m,2}G_1E_{m,2}G_4$;

(2) $\eta\sigma(E_{m,2}G_1E_{m,2}G_4E_{m,2}G_1E_{m,2}G_4)\eta^{-1} = E_{m,2}G_1E_{m,2}G_2E_{m,2}G_1E_{m,2}G_2$.

In fact,

$$\begin{aligned} \eta\sigma(E_{m,2}G_1E_{m,2}G_2)\eta^{-1} &= \eta\sigma(B_m A_m B_{m+1})\eta^{-1} = B_m * A_m A_m A_m B_m A_m B_m A_m B_m * B_m^{-1} \\ &= B_m A_m A_m A_m B_m A_m B_m A_m = E_{m,2}G_1E_{m,2}G_2E_{m,2}G_1E_{m,2}G_4E_{m,2}G_1E_{m,2}G_4; \\ \eta\sigma(E_{m,2}G_1E_{m,2}G_4E_{m,2}G_1E_{m,2}G_4)\eta^{-1} &= \eta\sigma(B_m A_m B_m A_m)\eta^{-1} \\ &= B_m * A_m A_m A_m B_m A_m A_m A_m B_m * B_m^{-1} \\ &= B_m A_m A_m A_m B_m A_m A_m A_m = E_{m,2}G_1E_{m,2}G_2E_{m,2}G_1E_{m,2}G_2. \end{aligned}$$

Step 3. We must prove the shift η is allowable. In fact, $G_0(E_{m,i}) = A_m$, then

$$\sigma(G_0(E_{m,i})) = \sigma(A_m) = A_m B_m = G_0(E_{m,i}) * \eta.$$

So under the substitution σ , there is a shift η . □

Example. Consider $E_{1,2} = aa$, then:

$$\begin{aligned} D_\infty &= \underbrace{ab}_{G_0(aa)} \underbrace{(a \ (a) \ a)}_A \underbrace{babab}_{B} \underbrace{(a \ (a) \ a)}_A \underbrace{b}_{C} \underbrace{(a \ (a) \ a)}_A \underbrace{b}_{C} \underbrace{(a \ (a) \ a)}_A \underbrace{babab}_{B} \underbrace{(a \ (a) \ a)}_A \underbrace{babab}_{B} \\ &\quad \underbrace{(a \ (a) \ a)}_A \underbrace{babab}_{B} \underbrace{(a \ (a) \ a)}_A \underbrace{b}_{C} \underbrace{(a \ (a) \ a)}_A \underbrace{b}_{C} \underbrace{(a \ (a) \ a)}_A \underbrace{babab}_{B} \underbrace{(a \ (a) \ a)}_A \underbrace{b}_{C} \underbrace{(a \ (a) \ a)}_A \\ &\quad \underbrace{b}_{C} \underbrace{(a \ (a) \ a)}_A \underbrace{babab}_{B} \underbrace{(a \ (a) \ a)}_A \underbrace{b}_{C} \underbrace{(a \ (a) \ a)}_A \underbrace{b}_{C} \underbrace{(a \ (a) \ a)}_A \underbrace{babab}_{B} \dots \end{aligned}$$

5. Gaps and Gap Sequence of an Arbitrary Word

In Section 3, we give the weak version of “uniqueness of envelope extension property”. In this section, we are going to give the strong version. These properties make “envelope word” quite special. Using them, we can extend the properties about gaps and gap sequence from envelope word to an arbitrary word in Doubling sequence D_∞ .

Theorem 5.1 (Uniqueness of envelope extension, strong version).

Let $Env(\omega) = E_{m,i}$, $i = 1, 2$, then ω_p has an unique envelope extension as

$$E_{m,i,p} = \mu_1(\omega) * \omega_p * \mu_2(\omega),$$

where $\mu_1(\omega)$ and $\mu_2(\omega)$ are constant words depending only on ω .

Proof. The key is to prove that $Env(\omega_p) = Env(\omega)_p$, i.e. $Env(\omega_p) = E_{m,i,p}$.

We write $G_j(Env(\omega))$ as G_j for short. We only need to prove that:

(a) ω occurs only once in $G_0Env(\omega)$;

(b) ω occurs only twice in $Env(\omega)G_jEnv(\omega)$.

Part 1. When $i = 1$, $E_{m,1} = A_m \delta_m^{-1}$. $G_0 = \varepsilon$, $G_1 = \delta_m$, $G_2 = \delta_{m-1} A_{m-1}^{-1}$. By the proof of Theorem 3.7, we have $\delta_{m-3} A_{m-3} \prec \omega$.

In this part, We write A_{m-3} (resp. B_{m-3}) as A (resp. B) for short.

(1) $G_0 E_{m,1} = E_{m,1}$, so ω occurs only once in $G_0 E_{m,1}$.

(2) $E_{m,1} G_1 E_{m,1} = A_m \delta_m^{-1} \delta_m A_m \delta_m^{-1} = A_m A_m \delta_m^{-1}$.

$$\begin{array}{c}
 [2]=2^{m-1} \quad [4]=13 * 2^{m-3} \\
 \downarrow \quad \downarrow \\
 E_{m,1,1} G_1 E_{m,1,2} = ABAAABABABAAABAB \delta_m^{-1} \\
 \uparrow \quad \uparrow \\
 [1]=3 * 2^{m-3} \quad [3]=3 * 2^{m-1}
 \end{array}$$

Fig. 5.1: The 4 possible positions of $\delta_{m-3} A_{m-3}$ in $E_{m,1,1} G_1 E_{m,1,2}$.

(a) When $\delta_{m-3} A_{m-3}$ occur at positions [1] and [2], $\omega \prec E_{m,1,1}$;

(b) When $\delta_{m-3} A_{m-3}$ occur at positions [3] and [4], $\omega \prec E_{m,1,2}$.

Since ω occurs once in $E_{m,1}$, ω occurs twice in $E_{m,1} G_1 E_{m,1}$.

(3) $E_{m,1} G_2 E_{m,1} = A_m \delta_m^{-1} \delta_{m-1} A_{m-1}^{-1} A_m \delta_m^{-1} = A_{m-1} A_m \delta_m^{-1}$.

$$\begin{array}{c}
 [2]=2^{m-1} \quad [4]=2^m \\
 \downarrow \quad \downarrow \\
 E_{m,1,2} G_2 E_{m,1,3} = ABAAABAAABAB \delta_m^{-1} \\
 \uparrow \quad \uparrow \\
 [1]=3 * 2^{m-3} \quad [3]=7 * 2^{m-3}
 \end{array}$$

Fig. 5.2: The 4 possible positions of $\delta_{m-3} A_{m-3}$ in $E_{m,1,2} G_2 E_{m,1,3}$.

(a) When $\delta_{m-3} A_{m-3}$ occur at positions [1] and [2], $\omega \prec E_{m,1,2}$;

(b) When $\delta_{m-3} A_{m-3}$ occur at positions [3] and [4], $\omega \prec E_{m,1,3}$.

Since ω occurs once in $E_{m,1}$, ω occurs twice in $E_{m,1} G_2 E_{m,1}$.

Part 2. When $i = 2$, $E_{m,2} = A_{m-1} A_m \delta_m^{-1}$.

$$G_0 = A_m, \quad G_1 = \delta_m A_m^{-1}, \quad G_2 = \delta_m B_{m+1}, \quad G_4 = \delta_m.$$

By the proof of Theorem 3.7, we have $\delta_{m-1} A_{m-1} \prec \omega$.

In this part, We denote A_{m-1} (resp. B_{m-1}) as A (resp. B) for short.

(1) $G_0 E_{m,2} = A_m A_{m-1} A_m \delta_m^{-1} = ABAAAB \delta_m^{-1}$.

The possible position of $\delta_{m-1} A_{m-1}$ in $G_0 E_{m,2}$ is $3 * 2^{m-1}$.

(2) $E_{m,2} G_1 E_{m,2} = A_{m-1} A_m \delta_m^{-1} \delta_m A_m^{-1} A_{m-1} A_m \delta_m^{-1} = B_m A_m \delta_m^{-1} = AA * AB \delta_m^{-1}$.

The possible positions of $\delta_{m-1} A_{m-1}$ in $E_{m,2} G_1 E_{m,2}$ are 2^{m-1} and 2^m .

(3) $E_{m,2} G_2 E_{m,2} = A_{m-1} A_m \delta_m^{-1} \delta_m B_{m+1} A_{m-1} A_m \delta_m^{-1} = AABABABAAAB$.

The possible positions of $\delta_{m-1} A_{m-1}$ in $E_{m,2} G_2 E_{m,2}$ are 2^{m-1} and 2^{m+3} .

(4) $E_{m,2} G_4 E_{m,2} = A_{m-1} A_m \delta_m^{-1} \delta_m A_{m-1} A_m \delta_m^{-1} = AABAAB$.

The possible positions of $\delta_{m-1} A_{m-1}$ in $E_{m,2} G_4 E_{m,2}$ are 2^{m-1} and 2^{m+2} .

Since ω occurs once in $E_{m,2}$, ω occurs once in $G_0 E_{m,2}$, twice in $E_{m,2} G_i E_{m,2}$, $i = 1, 2, 4$. \square

Theorem 5.2 (Gap sequence of an arbitrary word ω). Let $Env(\omega) = E_{m,i}$, $i = 1, 2$, then

(1) When $i = 1$, the gap sequence of ω is sequence Θ_1 on alphabet $\{G_1(\omega), G_2(\omega)\}$;

(2) When $i = 2$, the gap sequence of ω is sequence Θ_2 on alphabet $\{G_1(\omega), G_2(\omega), G_4(\omega)\}$.

Proposition 5.3 (Relation between $G_p(\omega)$ and $G_p(Env(\omega))$, $p \geq 1$).

Let $Env(\omega) = E_{m,i}$, $i = 1, 2$, then

$$G_p(\omega) = \mu_2(\omega) * G_p(E_{m,i}) * \mu_1(\omega), \quad p \geq 1.$$

Proof. The proof of the proposition is easy by the following diagram.

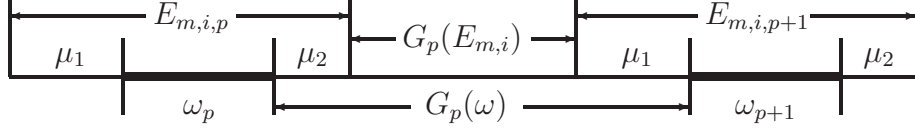


Fig. 5.3: The relation among ω_p , $E_{m,i,p}$, $G_p(\omega)$ and $G_p(E_{m,i})$.

□

Proposition 5.4 (Relation between $G_0(\omega)$ and $G_0(Env(\omega))$).

Let $Env(\omega) = E_{m,i}$, $i = 1, 2$, then

$$G_0(\omega) = G_0(E_{m,i}) * \mu_1(\omega).$$

Proof. The proof of the proposition is easy by the following diagram.

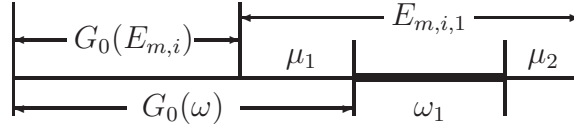


Fig. 5.4: The relation among ω_1 , $E_{m,i,1}$, $G_0(\omega)$ and $G_0(E_{m,i})$.

□

Using Theorem 5.1, Proposition 5.3 and 5.4, we can give the expressions of $G_p(\omega)$. The proofs are simple, while the expressions are complicated, we prefer to omit them.

6. The Positions of ω_p

The position of ω_p and ω_q ($p \neq q$) are distinct. In fact, ω_p should be regarded as two variables ω and p , where $\omega \prec D_\infty$ and $p \geq 1$. In this section, we will give the position of ω_p for all (ω, p) .

Definition 6.1. Let $N_j(x, p) := |\Theta_j[1, p]|_x$ be the number of letter x occurring in $\Theta_j[1, p]$, where $j = 1, 2$ and $p \geq 0$. When $j = 1$, letter x is an element of alphabet $\{a, b\}$; when $j = 2$, letter x is an element of alphabet $\{a, b, c\}$.

Example. Since $\Theta_1 = abbaaabbabbab \dots$, $N_1(a, 5) = |\Theta_1[1, 5]|_a = 3$, $N_1(b, 9) = |\Theta_1[1, 9]|_b = 4$.

Proposition 6.2 (Positions of $E_{m,i,p}$).

- (1) $L(E_{m,1}, p+1) = p * 2^m - N_1(b, p) * 2^{m-1} + 1$;
- (2) $L(E_{m,2}, p+1) = (3p+2) * 2^{m-1} - N_2(a, p) * 2^m + N_2(b, p) * 2^{m+1} + 1$.

Proof. In sequence D_∞ , there are p “ $E_{m,i}$ ”, one “ G_0 ” and gaps $G_j(\omega)$ ($1 \leq j \leq p$) ahead.

When $i = 1$, $E_{m,1} = A_m \delta_m^{-1}$, $G_0 = \varepsilon$, $G_1 = \delta_m$, $G_2 = \delta_{m-1} A_{m-1}^{-1}$.

$$\begin{aligned} L(E_{m,1}, p+1) - 1 &= p * |E_{m,1}| + |G_0| + N_1(a, p) * |G_1| + N_1(b, p) * |G_2| \\ &= p * |A_m \delta_m^{-1}| + |\varepsilon| + N_1(a, p) * |\delta_m| + N_1(b, p) * |\delta_{m-1} A_{m-1}^{-1}| \\ &= p * 2^m - p + N_1(a, p) + N_1(b, p) - N_1(b, p) * 2^{m-1} = p * 2^m - N_1(b, p) * 2^{m-1}. \end{aligned}$$

When $i = 2$, $E_{m,2} = A_{m-1} A_m \delta_m^{-1}$, $G_0 = A_m$, $G_1 = \delta_m A_m^{-1}$, $G_2 = \delta_m B_{m+1}$, $G_4 = \delta_m$.

$$\begin{aligned} L(E_{m,2}, p+1) - 1 &= p * |E_{m,2}| + |G_0| + N_2(a, p) * |G_1| + N_2(b, p) * |G_2| + N_2(c, p) * |G_4| \\ &= p * |A_{m-1} A_m \delta_m^{-1}| + |A_m| + N_2(a, p) * |\delta_m A_m^{-1}| + N_2(b, p) * |\delta_m B_{m+1}| + N_2(c, p) * |\delta_m| \\ &= 3p * 2^{m-1} - p + 2^m + N_2(a, p) - N_2(a, p) * 2^m + N_2(b, p) + N_2(b, p) * 2^{m+1} + N_2(c, p) \\ &= (3p+2) * 2^{m-1} - N_2(a, p) * 2^m + N_2(b, p) * 2^{m+1}. \end{aligned}$$

□

Proposition 6.3 (Position of ω_p).

Let $Env(\omega) = E_{m,i}$ and $Env(\omega) = \mu_1 * \omega * \mu_2$, then $L(\omega, p) = L(E_{m,i}, p) + |\mu_1|$, $i = 1, 2$.

7. Combinatorial Properties of Factors

As an application, we will give some combinatorial properties of factors in sequence D_∞ . Damanik [6] gave a description of the sets of palindromes and powers occurring in Doubling sequence. Using the properties of gaps and gap sequence of envelope words in Doubling sequence, we can get more explicit results.

7.1 Palindrome

Since envelope words $E_{m,i}$ are palindromes, we can determine all palindromes with envelope $Env(\omega) = E_{m,i}$. Using it, we can give a description of the set of palindromes.

Property 7.1. Let ω be a palindrome, $Env(\omega) = E_{m,i}$ and $E_{m,i} = \mu_1 \omega \mu_2$, then $|\mu_1| = |\mu_2|$.

- (1) When $i = 1$ and $m = 1$ (resp. 2), $|\mu_1| = 0$ (resp. $|\mu_1| = 0, 1$);
- (2) When $i = 1$ and $m \geq 3$, $0 \leq |\mu_1| < 3 * 2^{m-3}$;
- (3) When $i = 2$ and $m \geq 1$, $0 \leq |\mu_1| < 2^{m-1}$.

Proof. By Property 3.4, envelope words are palindromes. Suppose $|\mu_1| < |\mu_2|$, then

$$\overleftarrow{E_{m,i}} = \overleftarrow{\mu_1 \omega \mu_2} = \overleftarrow{\mu_2} \overleftarrow{\omega} \overleftarrow{\mu_1} = \overleftarrow{\mu_2} \omega \overleftarrow{\mu_1} = E_{m,i} = \mu_1 \omega \mu_2.$$

This means ω occurs in $E_{m,i}$ twice. It contradicts Theorem 3.7. So $|\mu_1| = |\mu_2|$.

Consider the range of $|\mu_1|$.

Case 1: $E_{m,1} = A_m \delta_m^{-1}$. For $m \geq 3$, suppose $|\mu_1| = 3 * 2^{m-3} + h$ where $h \geq 0$, then

$$\begin{aligned} & E_{m,1}[3 * 2^{m-3} + h + 1, 5 * 2^{m-3} - h - 1] \\ &= (A_{m-3} B_{m-3} A_{m-3} A_{m-3} A_{m-3} B_{m-3} A_{m-3} B_{m-3})[3 * 2^{m-3} + h + 1, 5 * 2^{m-3} - h - 1] \\ &= (A_{m-3} A_{m-3})[h + 1, 2^{m-2} - h - 1] \prec A_{m-2} \delta_{m-2}^{-1} = E_{m-2,1} \sqsubset E_{m,1}. \end{aligned}$$

When $|\mu_1| < 3 * 2^{m-3}$, $A_{m-3} A_{m-3} \prec \omega$. Since $A_{m-3} A_{m-3}$ isn't the factor of envelope word which less than $E_{m,1}$, then $0 \leq |\mu_1| < 3 * 2^{m-3}$ in this case.

When $m = 1$, $|\mu_1| = 0$; when $m = 2$, $|\mu_1| = 0$ or 1.

Case 2: $E_{m,2} = A_{m-1} A_m \delta_m^{-1}$. For $m \geq 1$, suppose $|\mu_1| = 2^{m-1} + h$ where $h \geq 0$, then

$$\begin{aligned} & E_{m,2}[2^{m-1} + h + 1, 2^m - h - 1] = (A_{m-1} A_{m-1} B_{m-1})[2^{m-1} + h + 1, 2^m - h - 1] \\ &= A_{m-1}[h + 1, 2^{m-1} - h - 1] \prec A_{m-1} \delta_{m-1}^{-1} = E_{m-1,1} \sqsubset E_{m,2}. \end{aligned}$$

When $|\mu_1| < 2^{m-1}$, $A_{m-1} \prec \omega$. Since A_{m-1} isn't the factor of envelope word which is less than $E_{m,2}$, $0 \leq |\mu_1| < 2^{m-1}$ in this case. \square

Corollary 7.2. All palindromes in sequence D_∞ are of odd length, except "aa".

Proof. Let ω be a palindrome. Since all envelope words are of odd length, except $E_{1,2} = aa$, and $|\omega| = |Env(\omega)| - 2 * |\mu_1|$, $|\omega|$ is odd. \square

Proposition 7.3. Let $P(n)$ be the number of palindromes of length n , then $P(1) = 2$, $P(2) = 1$, $P(3) = 3$. For $n \geq 3$, $P(2n - 2) = 0$ and

$$P(2n - 1) = \begin{cases} 3, & 3 * 2^n + 1 \leq n \leq 2^{n-2}; \\ 4, & 2^{n-2} + 1 \leq n \leq 3 * 2^{n+1}. \end{cases}$$

Proof. All palindromes with envelope $E_{1,1}$ or $E_{2,1}$ are $\{a, b, aba\}$.

All lengths of palindromes with envelope $E_{m+1,1}$ for $m \geq 2$ are

$$\{2^{m-1} + 1, 2^{m-1} + 3, \dots, 2^{m+1} - 3, 2^{m+1} - 1\}.$$

There are one palindrome of each length above.

All lengths of palindromes with envelope $E_{m,2}$ for $m \geq 1$ are

$$\{2^{m-1} + 1, 2^{m-1} + 3, \dots, 3 * 2^{m-1} - 3, 3 * 2^{m-1} - 1\}.$$

There are one palindrome of each length above too. So the proposition holds. \square

7.2 Power, Overlap and Separate Properties Between ω_p and ω_{p-1}

Property 7.4. Let $Env(\omega) = E_{m,1}$, then $G_0(\omega) \geq 0$, $G_1(\omega) > 0$ and

$$G_2(\omega) \begin{cases} > 0, & |\omega| < 2^{m-1}, \\ = 0, & |\omega| = 2^{m-1}, \\ < 0, & |\omega| > 2^{m-1}. \end{cases}$$

Property 7.5. Let $Env(\omega) = E_{m,2}$, then $G_0(\omega) > 0$, $G_1(\omega) < 0$, $G_2(\omega) > 0$, $G_4(\omega) > 0$.

By Theorem 4.2, 4.3 and Proposition 5.3, all factor $\omega \in D_\infty$ can be divided into four types according to the different lengths of gaps. We denote those types by T_j , $j = 1, 2, 3, 4$. Obviously, the disjoint union of sets T_j , $j = 1, 2, 3, 4$, consists of all factors in sequence D_∞ .

Definition 7.6 (Types). The sets T_j are defined as follow:

$$\begin{aligned} T_1 &= \{\omega \in D_\infty : Env(\omega) = E_{m,1}, |\omega| < 2^{m-1}\}; \\ T_2 &= \{\omega \in D_\infty : Env(\omega) = E_{m,1}, |\omega| = 2^{m-1}\}; \\ T_3 &= \{\omega \in D_\infty : Env(\omega) = E_{m,1}, |\omega| > 2^{m-1}\}; \\ T_4 &= \{\omega \in D_\infty : Env(\omega) = E_{m,2}\}. \end{aligned}$$

Corollary 7.7.

- (1) $\omega \in T_1 \iff G_0(\omega) \geq 0, G_1(\omega) > 0, G_2(\omega) > 0;$
- (2) $\omega \in T_2 \iff G_0(\omega) \geq 0, G_1(\omega) > 0, G_2(\omega) = 0;$
- (3) $\omega \in T_3 \iff G_0(\omega) \geq 0, G_1(\omega) > 0, G_2(\omega) < 0;$
- (4) $\omega \in T_4 \iff G_0(\omega) > 0, G_1(\omega) < 0, G_2(\omega) > 0, G_4(\omega) > 0.$

We will study some combinatorial properties such as adjacent, separated and overlapped of factors. From our knowledge, all previous studies on combinatorial only consider factor ω . We consider not only ω but also p , which is much more difficult.

Definition 7.8 (Properties). Let $\omega \in D_\infty$, $p \geq 1$, $i \geq 1$, define

$$\begin{aligned} \mathcal{P}_i &= \{(\omega, p) : G_j(\omega) = 0, j = p, p+1, \dots, p+i-1\} = \{(\omega, p) : \omega_p \cdots \omega_{p+i} \prec D_\infty\}; \\ \mathcal{S}_i &= \{(\omega, p) : G_j(\omega) > 0, j = p, p+1, \dots, p+i-1\}, \mathcal{S}_\infty = \{(\omega, p) : G_j(\omega) > 0, \forall j \geq 1\}; \\ \mathcal{O}_i &= \{(\omega, p) : G_j(\omega) < 0, j = p, p+1, \dots, p+i-1\}. \end{aligned}$$

Property 7.9 (Power Property). $\mathcal{P}_1 = (T_2, \Gamma_1(bb))$, $\mathcal{P}_2 = (T_2, \Gamma_1(bb))$, $\mathcal{P}_j = \emptyset$ for $j \geq 3$.

Property 7.10 (Separated Property).

$$\begin{aligned} \mathcal{S}_1 &= (T_1, \mathbb{N}) \cup (T_2 \cup T_3, \Gamma_1(aa)) \cup (T_4, \Gamma_2(b) \cup \Gamma_2(c)); \\ \mathcal{S}_2 &= (T_1, \mathbb{N}) \cup (T_2 \cup T_3, \Gamma_1(aa)); \\ \mathcal{S}_3 &= (T_1, \mathbb{N}) \cup (T_2 \cup T_3, \Gamma_1(aaa)); \\ \mathcal{S}_j &= (T_1, \mathbb{N}) \text{ for } j \geq 4 \text{ and } j = \infty. \end{aligned}$$

Property 7.11 (Overlapped Property).

$\mathcal{O}_1 = (T3, \Gamma_1(b)) \cup (T4, \Gamma_2(a))$, $\mathcal{O}_2 = (T3, \Gamma_1(bb))$, $\mathcal{O}_j = \emptyset$ for $j \geq 3$.

7.3 Squares and Cubes in Sequence D_∞

Brown, Rampersad, Shallit and Vasiga [4] described all the positions of squares occurring in Thue-Morse sequence M_∞ . They also counted the number of distinct squares beginning in $M_\infty[1, N]$ for $N \geq 1$. In this subsection, we are going to count the number of distinct squares (resp. cubes) beginning in $D_\infty[1, N]$ for $N \geq 1$. Factor ω_p is said to be begin in $D_\infty[1, N]$, if the first letter of ω_p occurs in $D_\infty[1, N]$.

Property 7.12 (Number of distinct squares and cubes).

Let $c(N)$ be the number of distinct squares beginning in $D_\infty[1, N]$. Then

$$c(N+1) - c(N) = \begin{cases} 1, & 2^m \leq N < 3 * 2^{m-1}; \\ 0, & 3 * 2^{m-1} \leq N < 2^{m+1}. \end{cases}$$

where $m \geq 1$. The number of distinct cubes beginning in $D_\infty[1, N]$ is $c(N)$ too.

Proof. Let $\omega\omega \prec D_\infty$, then $\omega \in T2$. Let $Env(\omega) = E_{m,1}$ and $E_{m,1} = \mu_1\omega\mu_2$. By Property 7.9, $\mathcal{P}_1 = (T2, \Gamma_1(b))$ and $\mathcal{P}_2 = (T2, \Gamma_1(bb))$. Since $\Theta_1[2] = b$ and $\Theta_1[2, 3] = bb$, by Theorem 5.2(1),

$$L(\omega\omega, 1) = L(\omega\omega\omega, 1) = L(\omega, 2) = |\mu_1| + L(E_{m,1}, 2) = |\mu_1| + 2^m + 1,$$

where $|\mu_1| + |\mu_2| = |E_{m,1}| - |\omega| = 2^{m-1} - 1$, so $0 \leq |\mu_1| \leq 2^{m-1} - 1$. The property holds. \square

Acknowledgments

The research is supported by the Grant NSFC No.11271223, No.11371210 and No.11326074.

References

- [1] J.M.Allouche, J.Shallit, Automatic sequences: Theory, applications, generalizations. Cambridge University Press, Cambridge, 2003.
- [2] J.Berstel, Mot de Fibonacci, Séminaire d'informatique théorique, L.I.T.P., Paris, Année 1980/1981, pp.57-78.
- [3] S.Brlek, Enumeration of factors in the Thue-Morse word, Disc. Appl. Math. 24 (1989) 83-96.
- [4] S.Brown, N.Rampersad, J.Shallit and T.Vasiga, Squares and overlaps in the Thus-Morse Sequence and some variants, RAIRO-inf. Theor. Appl. 40(2006) 473-484.
- [5] L.Balková, E.Pelantová, W.Steiner, Sequences with constant number of return words, Monatsh Math. 155 (2008) 251-263.
- [6] D.Damanik, Local symmetries in the period-doubling sequence, Discrete Appl. Math. 100(2000) 115-121.

- [7] F.Durand, A characterization of substitutive sequences using return words, Discrete Math. 179 (1998) 89-101.
- [8] Y.-K.Huang, Z.-Y.Wen, Gap Sequence of Factors of Fibonacci Sequence, arXiv:1404.4269.
- [9] M.Lothaire, Combinatorics on words, in: Encyclopedia of Mathematics and its applications, Vol.17, Addison-Wesley, Reading, MA, 1983.
- [10] M.Lothaire, Algebraic combinatorics on words, Cambridge Univ. Press, Cambridge, 2002.
- [11] C.Mauduit, Multiplicative properties of the Thue-Morse sequence, Period. Math. Hungar. 43(2001) 137-153.
- [12] J.J. Pansiot, The Morse sequence and iterated morphisms. Inform. Process. Lett. 12 (1981) 68-70.
- [13] Z.-X.Wen, Z.-Y.Wen, Some properties of the singular words of the Fibonacci word, European J. Combin. 15 (1994) 587-598.
- [14] L.Vuillon, A characterization of Sturmian words by return words, European J. Combin. 22 (2001) 263-275.